Nottingham University Business School

UNITED KINGDOM · CHINA · MALAYSIA

Analyzing the Customer reviews of hotels listed on Booking.com using Sentiment analysis to understand the Brand Perception

By

S M Sazid Islam

ID: 20350823

2021/22

A dissertation project presented as part of the degree of Masters in Business Analytics

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my research supervisor, Vanja Ljevar for her constant and excellent support, advice, and guidance throughout the development of this research work.

In addition, I want to dedicate this piece of work to my parents as a way of showing them how much I value their support throughout my life.

Finally, I want to show gratitude to the fellow mates of my Business Analytics course for being a very helpful part of my Master’s program experience. The journey would have been much more difficult without your motivation, assistance, and emotional support.

ABSTRACT

In today's world, the internet has become a powerhouse for a huge source of data in the form of user opinions, emotions, views, and arguments about various brands, products, topics, and events. The sentiment shared by an individual about a brand has a massive influence on other readers and viewers who are constantly scrolling the internet to search for answers to individual queries and needs. However, these data from the internet generally remain in unstructured form for which sentiment analysis is very much important to give the data a good structure and to interpret this huge compilation of emotions so that they can be classified and then segmented into different classes in order to understand the major emotions associated with a brand. Till now, many kinds of research have been conducted using different techniques of sentiment analysis to classify textual data. In this study, sentiment analysis and deep learning techniques have been merged to perform the sentiment analysis of reviews from the popular booking site called Booking.com. Machine learning has the ability to differentiate text and topic and that can be measured through model evaluations. The deep learning model is used because it is effective for their automatic learning capability. The study proceeds with a topic modeling to identify the highlighting topics used in negative reviews from a customer's point of view as they experience a stay in an accommodation. From topic modeling, this can be evaluated that topics like "room", "staff", "bathroom", "bed" and "breakfast" are more highlighted topics that guests focus on when expressing negative emotions in the reviews. In addition, sentiment analysis is conducted using sentiment scores and then classified into further classes of best, good, bad, and worst and it is discovered that most of the hotels listed in the Booking.com site is good as the majority of the reviews are in the good and best classes which is about 64% of total number of 515,738 reviews. Machine learning classification models like Random Forest, SVM, Naïve Bayes are used to train the dataset and accuracy scores are evaluated. Along with this, CNN model is also trained which shows the highest accuracy of 90% when compared with the traditional Machine learning algorithms. Hence, it can be stated that the deep learning Convolutional Neural Network (CNN) model is able to outperform the regular machine learning algorithms. Lastly, the top 10 best and worst hotels are also identified based on the customer's reviews. Booking.com should give suggestions to these worst hotels and come up with effective marketing plans to leverage them and the best hotels should also be announced in the website and application to make them more popular among the people.

Contents

List of Figures:

1. INTRODUCTION

The tourism industry has become one of the most uprising industries around the world. Day by day more and more people are spending more money in fancy holiday plans and for exploring the world by visiting different beautiful places. In addition, the advancement in transportation and roads and highways has paved the way for the emergence of a huge group of travelers. Hence, the need for suitable accommodation according to the budget and needs of the people has become a potential market. This is where the need for booking accommodations becomes an important aspect for people all over the world. User-generated reviews, comments, and reports about their travel experiences are becoming an increasingly important source of information with the advent of Web 2.0 (Habimana, et al., 2020). Specifically, for hotel booking, the reviews and experiences shared online are much more reliable and relevant for other consumers to follow than the promises made by hotels in promotions and advertisements. People prefer to read customer reviews more as it allows them to see from a consumer viewpoint and to observe the real picture. As a result, reviews and ratings are nowadays very much important for a hotel to compete in the market and maintain a strong customer base. This can only be achieved by giving customers the best living experience according to an individual's budget. Customer satisfaction has a direct impact on the brand image of the hotel, the higher the positivity a consumer expresses online in the form of review, the higher the number of customers will be willing to experience that service. A single positive and negative review can change the mind of many future customers who are looking for a place to stay for their next trip.

This need for booking a place for holidays and trips gave an ultimate opportunity for businesses to grow, more and more hotels are entering the business and making a profit. To bring these various hotels under one platform, a number of travels booking sites have evolved such as TripAdvisor, Booking.com, Expedia, Hotels.com, Kayak.com, Airbnb, etc. Among all these sites, Booking.com is one of the pioneers and remarkable websites which is one of the top choices of people when looking for accommodation because of its diversity of hotels and reliability of the information. Being a prominent travel website, the reviews posted in this site also plays an important role for the brand perception of each and every hotel in the list. Previously, many studies have been performed regarding sentiment analysis but this study is focused on analyzing the reviews using neural network. Moreover, Convolutional Neural Network (CNN) model is used for text classification because there has been a study where movie reviews are analyzed using this model (Rani & Kumar, 2019). Hotels located in various European countries are considered in this analysis.

The aim of the study is to make an in-depth sentiment analysis to understand the main topics that arise in bad reviews and to predict the sentiments correctly. In order to meet the aims, the data is at first preprocessed properly and then used for the classification. Also, traditional machine learning model and CNN models are used to assess the performance with this particular dataset.

In this study, a sentiment analysis has been conducted using deep learning methods and machine learning algorithms of the reviews of Booking.com. As well as, topic modeling is executed on the negative reviews to extract the main topics associated with negative emotions. Each and every review is given a score based on the sentiment and then the reviews are evaluated and classified into four classes: best, good, bad and worst reviews. Sentiment analysis is the process of classifying the emotions or views from texts into categories such as positive, and negative.

1.1. Research Question

The research questions define the hypothesis of the study, in this study we are working mainly with 2 research questions:

- What are the main topics that arise in bad reviews of Booking.com?
- How well can we predict positive and negative reviews?

This study is performed to identify the main areas where hotels and accommodations should focus while serving the customers so that it is a win-win situation at both ends, the customer gets what they are expecting and the hotel also gets rewarded with a loyal customer. This will help to reduce the gap between customers and a brand enabling them to have a long-term strong relationship with brands while brands having a good brand perception.

1.2. Objective

The objective of the study is to analyze the reviews and to have an in-depth classification of reviews so that the actual scenario of the reviews can be identified and based on that the brand perception of the company can be evaluated. This will allow Booking.com to observe a clearer image of the reviews and will be able to have a more focused sight of the problems faced by customers. Hence, the poor and best performing hotels are also identified.

2. LITERATURE REVIEW

The topic of the research is to perform a sentiment analysis on the reviews of Booking.com and the brand perception is evaluated. In order to proceed with the research, a thorough review of the literatures regarding this topic has been conducted prior to the start of this study. Through study it is examined that there has been a number of ways by which sentiment analysis has been done on textual reviews like using different techniques of Machine learning, using questionnaire forms to assess the emotions of the customers. In this study deep learning neural network is used to perform the sentiment analysis and classify them into classes based on scores given to each review. The further details of the literatures reviewed are described in the next paragraphs in this section.

In order to distinguish between positive and negative reviews, it is vital to analyze previous studies which allowed us to work on this study. This section includes the findings of the previous studies which lead to this research. Reviews and ratings play a significant role in shaping consumer perception of brands in today's competitive digital market. The rating and reviews establish the standard and the consumer's opinion of the product or service. Customer evaluations have been found to have a greater impact on consumers than specific marketing and promotions, according to studies. On microblogging websites today, users share their thoughts on a variety of subjects, including products, media, organizations, services received, etc. However, for a business to develop and get better over time, a true and transparent representation of consumer feedback is crucial. In estimation structures and perception mining, sentiment analysis is key. Customer reviews are an integral part of websites like Zomato, Yelp, Swiggy, and Amazon where consumers can publish their opinions on businesses, products, and services using unstructured text reviews and star ratings, typically out of five (Asghar, 2016). Similarly, booking.com is a customer-oriented company. They assist in giving customers access to the top hotels based on their needs. Studies based on customer satisfaction and dissatisfaction have been conducted by Berezina (K. Berezina A. Bilgihan, 2016). According to Harrison-Walker (Harrison-Walker, 2001), most hotels ought to take client complaints given the advantages they gain by creating websites, phone centers, and live chat services. One of the biggest online travel organizations in the world is Booking.com, where millions of people search for lodging among a variety of properties such as hotels, apartments, guest homes, and more. Consumers' expectations can be used to assess the quality of hotel products based on the following factors: ability to acquaint customers with hotel employees, perception toward customers, meeting customers’ expectations, and special offers or value that customers can actually experience (M. Starkov, 2018).

Big data as a term can be perceived in different manners based on the people, in simple terms it is just a collection of information. Big data analytics is a broad notion that has been characterized as a method for comprehending situations by studying, processing, discovering, and exhibiting the outputs with the advantages of cost reduction and prompt execution (Zhang & Kim, 2021). The importance of big data is inevitable in the hotel and tourism industry in today's world because of the huge benefits that can be achieved using the proper use of big data. The main and foremost goal of a hotel or any accommodation is to make the customers feel that they are having the best accommodation experience according to their budget. Massive volumes of data are produced during a customer's journey with a hotel, and with a well-planned Big Data strategy, hoteliers can now further enhance their high-quality service (Datafloq, 2014). The proper use of big data can lead to increased customer satisfaction, and develop personalized marketing campaigns, and offers so that the right visitors book the right hotel at the appropriate time and rate. Using deep customer insights, it is possible to give the guests an unforgettable experience. It helps a lot in forecasting the likes and dislikes of a consumer which ultimately helps the hotels to offer the exact things that a customer was looking for. Customer categorization is another important factor that can be conducted using the use of big data, this is a factor that helps in revenue generation of a business very impactfully as the better a business can recognize its customer base, the stronger relationship they can build.

Prior research has revealed that nearly 95% of people go for internet hotel evaluations written by real people before making a reservation (M. Ady, 2015). This demonstrates the value of internet evaluations and their veracity because there will otherwise be a significant disconnect between expectations and reality. There are academic studies on the customer reviews of Booking.com, previously, Isidoros Perikos, Argyro Tsirtsi, Konstantinos Kovas, Foteini Grivokostopoulou, Ioannis Daramouskas, and Ioannis Hatzilygeroudis have analyzed and visualized visitor reviews of hotels as expressed at booking.com portal which performs an automatic aspect-based analysis of reviews and which follows an aspect-based approach using latent Dirichlet allocation (LDA) to model topic opinions. Their study has focused on the important criteria of the services and the factors that a user looks for in a review (Perikos, et al., 2018). In addition, some natural language processing approaches are used to analyze textual reviews, specifying the dependency on a sentence level and assisting in understanding users' opinions. Their performance is evaluated by identifying the polarity of the users' opinions. Finally, they have trained a number of classifiers using various feature sets derived from textual reviews and integrated in ensemble schemas.

The study's findings demonstrate that dependencies and other factors helped the classification process perform better and that ensemble schemas outperform individual classifiers in a consistent manner.

There have been studies on hotel products and services where the travel purposes have been examined using customer reviews. In that study, 4 key attributes that drive customer satisfaction and 5 key attributes that drive customer dissatisfaction have been identified. In addition, their findings revealed that hotels with various star levels differ greatly in terms of client satisfaction and discontent with aspects of goods and services (Xu, et al., 2017). Text mining and regression approaches were used in that study to provide insights for the businesses in order to use customer reviews for brand perception (Xu, et al., 2017). In a different study, it has been analyzed that customers tend to have both positive and negative sentiments towards higher star level and editor-recommended hotels (Xu, 2020). Positive and negative textual components have been shown to differ from one another, and in comparison, the quantity of negative textual elements is higher and their contents are more particular (Xu, 2020). In another previous study, the relevance of core attributes of products and services of hotels was demonstrated using Latent Semantic Analysis (LSA) and statistical tests and the findings include significant effects of attributes like the physical setting of hotels, staff, location, the value had different impacts on guest satisfaction and dissatisfaction with the change in star levels of the hotels (Xu, 2018).

In another publication, the study shows the distribution of online ratings to bring hotel class into the picture, the distribution of hotel reviews by hotel class is contrasted to determine whether it varies across classes. This study has examined the distribution of hotel online ratings based on skewness and kurtosis. The data is cleaned and the analysis is conducted using a web crawler developed using Python programming language, several data science techniques and tools including data understanding are used in this study. As well as, non-parametric kernel density estimators have been used to make the analysis properly (M. Mariani & Borghi, 2017). The objective of this study by Marcello M. Mariani was to examine quantitatively if the scores are somewhat inflated towards the higher end of the scale for an entire population of hotels, they focused on London only as it ranks third among the top 100 city destinations worldwide (Euromonitor International, n.d.). However, they did not use any machine learning algorithms in this study to make further analysis.

In a different study, using probabilistic latent semantic analysis, the Probability Latent Semantic Analysis (PLSA) approach is utilized to detect sentiment in comment titles on Booking.com. Using terms with several meanings that are grouped together in the same words and used to identify client testimonials, PLSA develops a document context. Additionally, the probability of words and documents from customer testimonials was calculated using the PLSA approach. They also used the data crawling method using WebHarvy software which is done for processing and then after preprocessing the data through cleaning steps, the PLSA method is used to determine the good or bad testimonials and also to improve the accuracy of undetectable words on SentiWordNet. This paper also focuses on the measurement of similarity of documents which was done to know the categories in positive or negative words using Cosine Similarity (Khotimah & Riyanarto , 2018). However, the study was more based on detecting the sentiment of the comments and brand perception of booking.com was not focused.

In a different publication, a study of roughly 150 machine learning applications with effective consumer interactions conducted by numerous teams at booking.com was employed. The study's goal was to demonstrate how machine learning may have a significant influence in an industrial setting with significant financial gains. They focused on the challenges like the high stakes of disengagement with the platform when a customer has a negative experience, infinitesimal queries, the complexity of booking, etc. For this study, they developed a machine learning model that can contribute to solving the business case. The value of a certain measure that depends on the job is estimated using cross-validation of various methods (classification, regression, ranking). One of the important aspects was the value a model brings to customers and the business of booking.com. The models used in this study comprise the traveler context model, Item space navigation models, User interface optimization models, content curation, and content augmentation. The models can be used by various teams to develop their products and learn from the customer's point of view, however, it has been found via research that achieving the actual business effect is challenging. Their model can be utilized for many and very varied products in significantly different situations. Following a comparison of the models and an analysis of the connections, six lessons were developed from 150 successful machine learning applications in large-scale e-commerce (Bernardi, et al., 2019).

Online evaluations from previous customers are an important source for tourism research since they can be used to assess the level of service provided and the level of consumer satisfaction at various hotels and destinations (Rodrı´guez Dı´az & F Espino Rodrı´guez, 2018). Moreover, customers tend to express the most in the reviews so a real image of the scenario can be understood. Several studies have examined how internet reviews affect consumers' opinions, purchasing patterns, and service providers' performance. Manuel Rodriguez and Tomas F Espino Rodriguez conducted a study on establishing the reliability and validity of online reputation databases for lodging: Booking.com, TripAdvisor, and HolidayCheck. They looked at three of the most significant online reputation websites in tourism in order to establish the reliability and validity of the scales used in customer reviews. The study's primary goals were to assess the validity and reliability of the scales of variables utilized by websites, and its secondary goal was to examine the feasibility of doing so by introducing a nonparametric test to gauge validity. The variables are grouped in two groups for study purposes and the variables evaluated for booking.com mentioned are staff, comfort, facilities, cleanliness, location, Wi-Fi, and value for money. Based on these variables they conducted the Kruskal-Wallis test and evaluated the reliability and validity based on the scores (Rodrı´guez Dı´az & F Espino Rodrı´guez, 2018).

A brand's image and reputation are very much important because it creates the perception of the brand to the stakeholders and the external public. A previous study on the branding of booking.com has been done by the use of a hybrid method using both qualitative and quantitative analysis which examines the brand personality traits contained in hotel and users' content and using the emotional tone used on different platforms (Borges-Tiago, et al., 2021). The difference between brand identity and brand image across the various social media platforms was measured in this study using BPS. Additionally, sentiment analysis has been utilized to examine the consumers' emotional state because the manner that opinion strength is determined is tied to the intensity of specific emotions. The analysis also includes a study through the application of a predefined dictionary using the brand personality dictionary and sentiment dictionary. The focus of the study was on hotels located on an island destination to keep the context of the information search remains constant (Borges-Tiago, et al., 2021). Moreover, content analysis was carried out on various aspects to find the answers to the issues developed based on the study (Borges-Tiago, et al., 2021). Word of mouth (WOM) is also a very triggering factor for influencing a brand perception remarkably, in today's fast-paced digital world people prefer and rely on Electronic Word of Mouth (eWOM).

In a previous study, hotel managers' responses to online hotel reviews have been examined, and explained the impact of the responses on ratings of hotels (Lee & Blum, 2015). From the study, this was discovered that managers mostly responded to the positive reviews only with a very basic thank note rather than appreciating the guests and welcoming them for a further revisit. The study also states that managers should take the reviews more seriously and respond properly to both positive and negative reviews as it helps in creating more positive eWOM resulting in more sales. In another paper, unigram features with two types of information (frequency and TF-IDF-Term Frequency-Inverse document frequency) are being used to identify the polarity classification of documents (SHI & LI, 2011). The study shows that TF-IDF is a factor that is more effective and important than frequency. To study the research on textual reviews, a study on performance evaluation of sentiment classification has been conducted using two machine learning classifiers (Naïve Bayes and SVM) as well as SentiWordNet is also used in that research to make the sentiment classification of movie reviews (Singh, et al., 2013). A different concept has been studied from another research, the findings of the study mention that the higher the hotel category the higher the price set by the hotel (Fuentes, 2016). The particular study was conducted using both booking.com and TripAdvisor data. It also states that the overall quality of hotels can be inferred from the stars of ratings which customers use for hotel selection (Fuentes, 2016).

Based on previous studies and the research purposes, the big data analytics performed in this research includes a sentiment analysis approach along with a deep learning model implementation which can be used to distinguish between good and bad reviews. The difference between ratings and reviews is that ratings are structured while online textual reviews are unstructured user-generated content. Hence, online textual reviews can highlight the customer experience and thoughts in a more detailed manner (Wang, et al., 2019).

In the aforementioned studies, it can be seen that most of the studies have been done on sentiment analysis, and brand perception, using qualitative and quantitative techniques like creating questionnaire forms and conducting surveys of consumers, and learning about customer perceptions. However, there is a gap between brand review prediction using particular user sentiments with machine learning and deep learning algorithms. In this study, the textual reviews are classified into more specific sentiments to understand the emotions of the customers better and to pinpoint the factors causing the negative reviews. The study has been conducted using Convolutional Neural Network (CNN) to do sentiment analysis because previously there has been studies where text classification is performed using CNN, one of the references of the journal article is “Deep Learning Based Sentiment Analysis Using Convolutional Neural Network” by Sujata Rani, Parteek Kumar. In this study, they performed sentiment analysis using CNN on Hindi movie reviews compiled from different online newspapers and websites. The CNN model used was trained using 50% of the dataset and tested on 50% of the data (Rani & Kumar, 2019). Similarly, in this study, CNN is used to make an in-depth sentiment analysis of customer reviews of popular hotel booking site, Booking.com.

3. METHODOLOGY

The methodology consists of two parts. In the first part details about the dataset used and the brand studied are described and the second part consists of the methods used to analyze the dataset and the steps involved in getting the result is explained. The review of the guests who had used the service of booking.com is taken from Kaggle which has authentic and reliable data collected from booking.com directly. Kaggle is a reliable platform for data collection.

3.1. Data

The data file contains 515,738 customer reviews and the score given by customers on 1493 hotels across Europe. In addition, the geographical location is also mentioned. There are 17 fields in the dataset. Among the 17 features the study has been performed based on the textual reviews only for which the Negative and Positive review features are only considered to conduct the sentiment analysis as these are the main focus of the study. The sentiment analysis is the method of analyzing the emotions expressed in a text. It is very much important because if these emotions can be unearthed properly then the characteristics of the customer group can be easily predicted.

The data is large and informative and has been scraped directly from booking.com for which it is very authentic to use for analysis purpose and the results are expected to be more accurate.

Booking.com is one of the world's leading digital travel companies which is founded in 1996 in Amsterdam. The mission of the company is "make it easier for everyone to experience the world". Accommodations all over the world can connect with people around the world through booking.com. Booking.com made its services available in 43 different languages with its vast accommodation listings of 28 million with listings other than hotels is 6.2 million, places to stay like homes, apartments and other unique places. It helped the people by providing a place to stay anywhere and at any time and assisted with 24/7 customer support.

3.2. Method

Research Design:

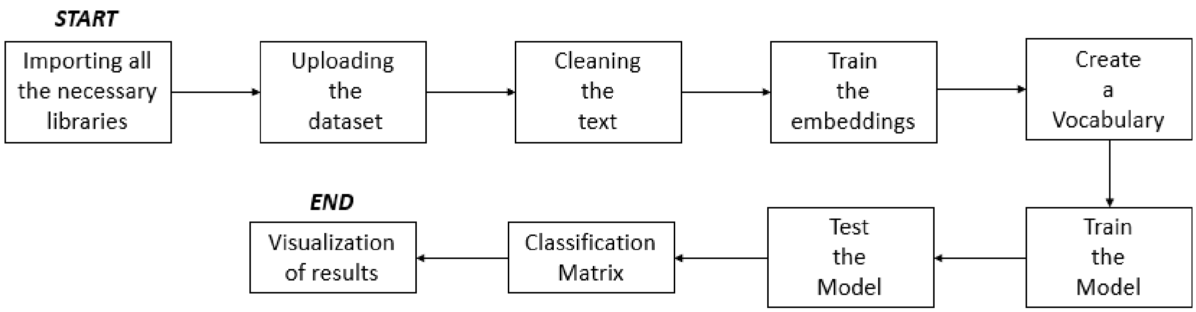

Figure 1: Research Methodology Main Phases

The above flow diagram shows the steps involved in the study to perform the sentiment analysis. The coding involved in this study has been performed using Google Colab which uses Python language programming.

This part is about the methods and steps involved in analyzing the reviews and reaching the objective of the study. The objective of this study is to perform a sentiment analysis task with the use of neural network on the reviews text. Positivity and negativity scores are extracted out of the reviews expressed in the range from 0 to 1. Where 0 denotes extremely negative and 1 is extremely positive. It is a method to rank the reviews of booking.com using sentiment analysis. The details of the different stages involved in methodology is described sequentially:

A. Data Pre-processing:

The first step after importing the dataset is cleaning the data properly to make it suitable for the analysis to get rid of any noise from the data. The textual reviews are cleaned by lowering the alphabets to lower case, by replacing words like “isn’t” with “is not”, the punctuations are removed from the text, the parenthesis, forward slash, backward slash, dashes, hashtags, and symbols are removed from the text. These steps are very important because it determines the efficiency of the whole analysis.

B. Topic Modeling:

It is the method of identifying a group of words or topics from a document or section of a file that show the prominent or most talked topics in the file. It helps to understand the most used topics and also it can add new dimension to a study by discovering hidden topical pattern that are present in the data. There are 3 ways to perform topic modeling, Non-Negative Matrix Factorization (NMF), Latent

Direct Allocation (LDA) and Latent Semantic Analysis (LSA) (Li, 2018). In this study, NMF and LDA are used to perform topic modeling. NMF is a statistical method that has the ability to reduce the dimension of the input text. One of its character is that it gives less weight to the words with less coherence according to the topic (Manthiramoorthi, n.d.). Term Frequency - Inverse Document Frequency (TF-IDF) is used for NMF, this is used to evaluate a term's significance within a document in relation to the document.

On the other hand, LDA is used to explore unstructured datasets by using the relationships that exist between the words of data. It uses Dirichlet distributions in its algorithm, two distributions are used, one is over topics and the other one is over words. Topic modeling is done here on the negative reviews to identify the topics customers talk more about in the negative reviews. The most used words are evaluated and the frequency of the words is visualized so that booking.com can suggest the areas about which the accommodations should be more concerned.

C. Support Vector Machine (SVM):

SVM is a simple supervised machine learning algorithm used for classification and regression purposes, the main character of SVM is that it creates a boundary between two classes of data in order to classify (Stecanella, 2017). It works on the geometric interpretation of the problems. SVM is used as a classification model because it can handle the high dimensional problem well as it is independent of dimensions and text being a high dimension problem. The goal of SVM is to maximize the margin between the support vectors and the hyperplane.

D. Random Forest Classifier:

It is a supervised learning algorithm used in solving regression and classification problems. It is a powerful and easily adaptable supervised machine learning algorithm that grows and combines a number of decision trees and creates a forest (Yiu, 2019). It merges them together to gain absolute and stable value. It is used in this study because Random Forest can deal with high dimensional data with noise in text classification.

E. Naïve Bayes Classifier:

The Naïve Bayes is a well-known machine learning algorithm that operates on the basis of probability. This model provides a more accurate and valid result than other algorithms or models since it does not consider the unnecessary elements of a given dataset to forecast the result (Gandhi, 2018). It is a classifier that employs every feature in the feature vector and analyses each one as an equal identity.

It is used in this study since it is a classification problem and it has the ability to predict properly with high accuracy.

F. Convolution Neural Network (CNN):

To perform the sentiment analysis of our dataset we have used CNN model. CNN is used as they have proven to be successful at text classification problems. CNN is a deep learning neural network developed for processing organized arrays of data, such as images (Patel, 2018). It is also used for text classification because it has the ability of natural language processing. It consists of a number of layers which tries to identify a pattern or useful information from the dataset. In order to provide a non-linear relationship for the output an activation function is used. In this study, ReLu is used as the activation function.

G. The Multilayer perceptron:

The multilayer perceptron is a deep learning method that is a feedforward artificial neural network that has the capability to generate a set of outputs from a set of inputs. It has several layers of input nodes connected as a directed graph between the input and output layers, this means that the signal path through the nodes only moves in one direction (Abirami & Chitra, 2020). Whereas, each node has a nonlinear activation function except for the input nodes. There are at least three layers in a Multilayer perceptron model: input, output, and hidden layers, while the size and number of the hidden layers can be changed according to the analysis need. It uses a supervised learning technique called backpropagation for training (Bento, 2021). Training data is propagated to the Multilayer Perceptron through input layers, it passes through the hidden layers if any forwarding outputs of activation functions to the next layer. Finally, the output is generated at the output layer by applying activation functions. By comparing predicted output and actual output the error is calculated.

In this study, neural networks are created using Keras Sequential Model which arranges the Keras layers in sequential order. The data flow takes place only in one direction and this model is limited to just one input tensor and one output tensor. For the purpose of analysis, the layers created here are dense which is the 1st layer of the neural network which is the most common and frequently used layer which performs a matrix-vector multiplication. The next layer is used in flatten layer which is used to flatten the input shape, this is used here to flatten all the reviews to one plane (ML, 2017). Embedding layer is also used which is used for language modeling, this is used because word embeddings is required for the study. Conv1D is another layer that creates a convolution kernel that is convolved with the layer input over a single temporal dimension to produce a tensor of outputs (Solanki, 2022). GlobalMaxPooling1D layer is also used to downsample the input representation by considering the maximum value.

H. Training the Multilayer perceptron model:

In order to train the multilayer perceptron model, the text of the reviews needs to be extracted and truth values need to be assigned. Before assigning the truth values, the data is shuffled to keep the model general and to avoid overfitting also it prevents any bias during the training. Since the positive and negative reviews are given in the dataset, new columns are created with the truth values. Two other datasets are created, one containing all the reviews with truth values and the other with all reviews to train embeddings. Next to train the embeddings the dataset with all reviews is used, for embedding Word2Vec skip-gram model is used which is a simple architecture for computing word embeddings that uses the technique of predicting the surrounding words by the use of a central word rather than using surrounding words to predict the center word (Mumbi, 2021). Word embedding maps words from the lexicon to vectors of real numbers using language modeling and feature learning techniques. The Word2Vec skip-gram is used in this study because it performs a fake task of predicting the words that will be closest to another word given that word's appearance in a sentence (Kulshrestha, 2019). The benefit of using Word2Vec is that it can produce very high-quality word embeddings very efficiently. The purpose of this fictitious task is to train the network in preparation for the extraction of the weight’s matrix, where each row will contain the vectoral representation of a word. In order to have a 128-dimensional representation of the words, the network is trained with 128 neurons in the hidden layer.

I. Creating a vocabulary:

It is essential to define a vocabulary of known words when using bag of words or embedding model. Using Keras Tokenizer a vocabulary is created of all words that appear in the reviews. The tokenizer is used here to convert each text into a vector that has a coefficient for each token in binary values. There are many methods of Keras tokenizer such as fit_on_texts, texts_to_sequences, texts_to_matrix and sequences_to_matrix. In this study, fit_on_texts method is used to update the internal vocabulary for the reviews.

J. Train Embedding layer:

After training the Word2Vec model and creating the vocabulary, the vectorial representation of words is combined to create an embedding matrix. When encoding text, a word embedding represents each word in the vocabulary as a real valued vector in a high-dimensional space. The vectors are learned in such a manner that words with similar meanings will be represented similarly in the vector space. This is used to initialize the weights of the embeddings layer in the perceptron and it is checked that there are no words without pre-trained embedding.

K. Training and Testing the model:

The text is then preprocessed and using the tokenizer, each word in the review is mapped with a one hot representation. Before splitting into training and test sets the reviews are padded with zeros to make them all of the same size and the size of each review considered is 40 words as it has been analyzed that the average length of reviews is 33 words. Then the text is divided into training and test sets. The hold-out method is used to split the data into multiple parts and then one part is used for training the model and the rest is used for validating and testing it. In the holdout method, maximum data is allocated to the training set and the remaining is used for testing. Test size is considered 33% which is used for prediction and training size is considered the rest 67%. The test and training size are set by checking the accuracy of the model by using different ratios. This ratio is chosen because the accuracy is good with this ratio. Next, the model is trained with 5 epochs to get a good accuracy of the model. Epochs are the number of iterations the whole dataset moves across the neural network. The model is then tested and test accuracy of 90% is achieved. To have a better visualization of classification results, a confusion matrix is plotted.

The scores for positivity and negativity of the sentiments are then extracted which are used as numerical features. Finally, based on the scores the reviews are divided into 4 classes:

- Best: final_score >= 0.7
- Good: final_score < 0.7 AND final_score >= 0.5
- Bad : final_score < 0.5 AND final_score >= 0.3
- Worst : final_score < 0.3

The number of reviews in each class is examined and the best and worst hotels based on the scoring system of sentiments is evaluated.

4. RESULTS

This section includes the outcomes of the analysis which are presented in graphs and charts. This study uses a data source from an online data science platform called Kaggle. The data consists of 515,738 reviews of guests who used booking.com to book hotels or accommodations in different countries across Europe at 1493 hotels. All the reviews are in English language with 17 features about accommodations and reviewers. After preprocessing the data properly, the data is free from any brackets, irregular signs, symbols and punctuations. This is done to make it suitable for the neural network.

The positive and negative scores of the sentiments are presented in graphs.

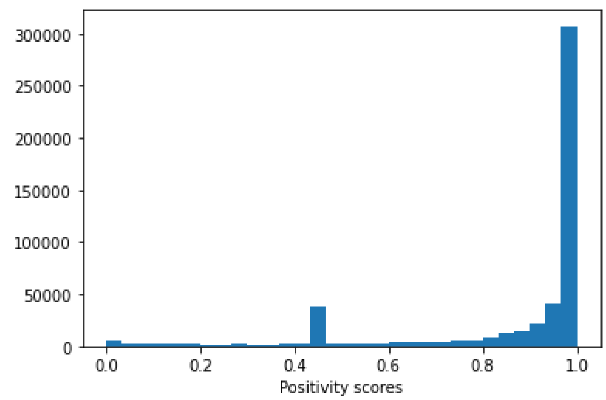

Figure 2: The distribution of positivity scores of the reviews

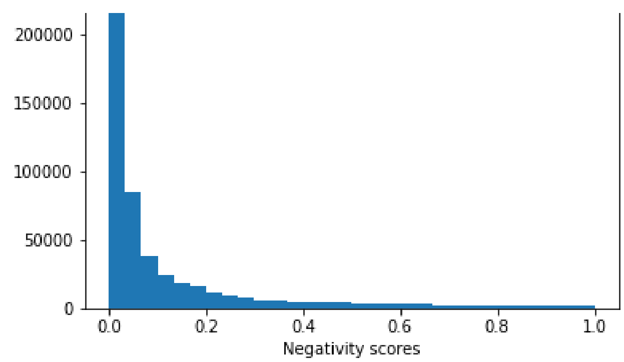

Figure 3: The distribution of negativity scores of the reviews

The positive and negative scores show the variation of scores of the sentiments of the texts. The text reviews are scored using the Multilayer perceptron model, each review is given a score based on the positivity and negativity of the review. The distribution of the scores of the texts are shown in the above graphs. Scores near to 1 are highly positive and scores near to 0 are highly negative reviews.

After analysis, two problems are identified that few data are misclassified, that is some negative reviews appears to have a positive score and some positive reviews appear to have a negative score. Another issue is there are guests who gave two reviews, and shared both positive and negative reviews. In this case, the review with higher score is considered.

4.1. Final visualization of scores

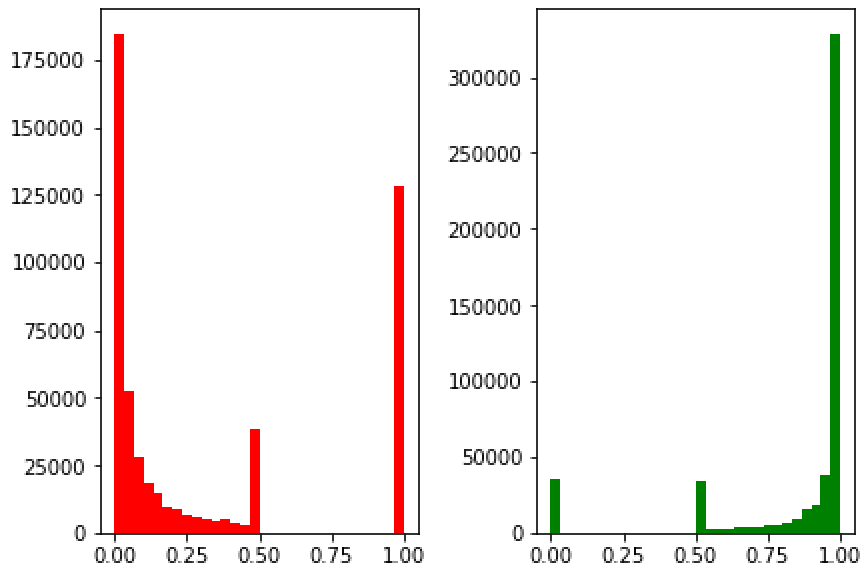

Figure 4: The final distribution of scores:

The above graph shows that most of the positive reviews are highly positive, ranging in between 0.5 and 1.00 while most of the negative reviews are in between 0.0 and 0.5.

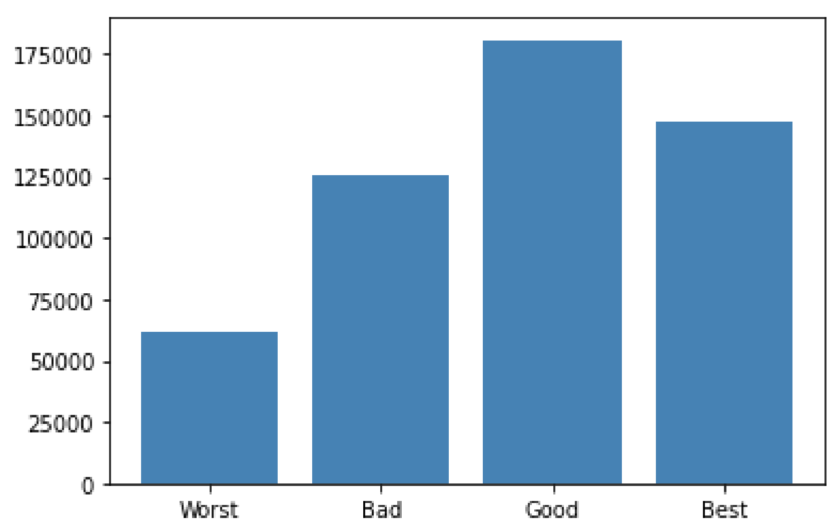

Figure 5: Dividing the reviews into 4 classes

The above figure shows the number of reviews in each class ranging from Worst to Best. After analysis, the scores of the reviews are distributed as follows:

Number of Best reviews: 147747

Number of Good reviews: 180787

Number of Bad reviews: 125822

Number of Worst reviews: 61382

The distribution of different classes based on analysis:

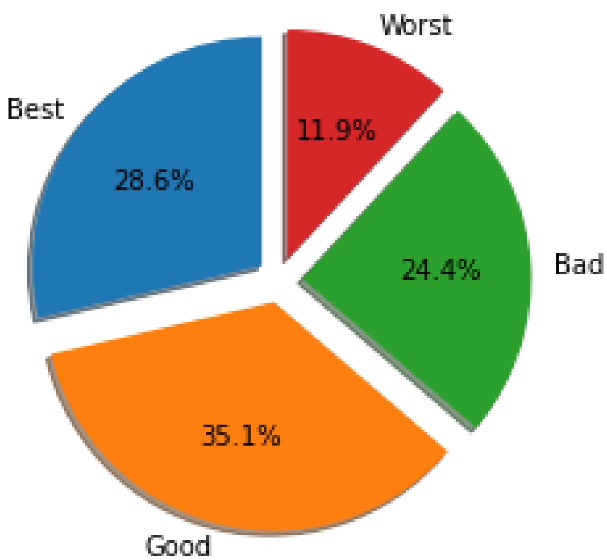

Figure 6: Pie Chart to show the distribution of reviews

Among the total number of 515,738 reviews, 35.1% is considered as Good, 28.6% is Best, 24.4% is Bad and 11.9% is Worst. The variation shows that more than 50% of the reviews are positive and guests have a very good impression of the hotels listed in Booking.com.

4.2. Topic Modeling

	unigram	count
0	room	176026
1	negative	129447
2	hotel	74709
3	breakfast	58478
4	small	49880
5	staff	39512
6	rooms	34802
7	bed	29828
8	bit	27546
9	bathroom	26585

Figure 7: Topic Modeling unigram

After applying topic modeling to the negative reviews of Booking.com, the most used words by the reviewers and their frequency can be visualized.

4.3. Model Evaluation

A number of classification models has been used in this study to train the dataset and identify the best model which can predict accurately. The models used are Support Vector Machine (SVM), Random Forest, Naïve Bayes and CNN.

SVM: After training the dataset using this model, the accuracy is found to be 55% which shows that the mode is performing not that well with this particular hotel reviews dataset and the recall is 56% which is also an average score.

Random Forest: When the dataset is trained using this model, the accuracy is 72% which shows that the model is predicting moderately, not that well but higher than SVM. The recall is 74% which is greater than SVM and this implies that the model can identify the actual positives out of all the positives better than SVM. So Random Forest can predict the data more accurately than SVM.

Naïve Bayes Classifier: The accuracy of this model is 54% which is lower than the Random Forest classifier. However, the recall value is far higher than Random Forest with a value of 87%. This indicates that even though the accuracy of the model is low but the model can accurately predict the true positives and true negatives from the dataset. The ratio of TP/(TP+FN) is high, which means that the value of False negative is low.

CNN: When the data is trained using the CNN model, the accuracy is found to be 90% which is the highest among the other classifiers used with a recall value of 90%. This shows that the value of false negative is very low for this model for which it has a high recall value. This clearly indicates that it has very high accuracy and a high ability to predict the true positives and negatives. However, high accuracy can lead to overfitting but in text classification overfitting do not have high impact.

Classification report table to show the variation in performance for different classifiers:

CLASSIFIER	ACCURACY	RECALL
SVM	55%	56%
Random Forest	72%	74%
Naïve Bayes	54%	87%
CNN	90%	90%

4.4. Final selection of model

Based on the classification reports of all the models, CNN is considered as the best and suitable model for this study because this is a text classification study and here high accuracy will not impact remarkably on the dataset. Hence, impact of overfitting will not be a problem for this analysis. Moreover, the number of predicting true positive (TP) and true negative (TN) is highest using this model compared to the other models for which it has a high recall value.

The Best 10 hotels according to the reviewers:

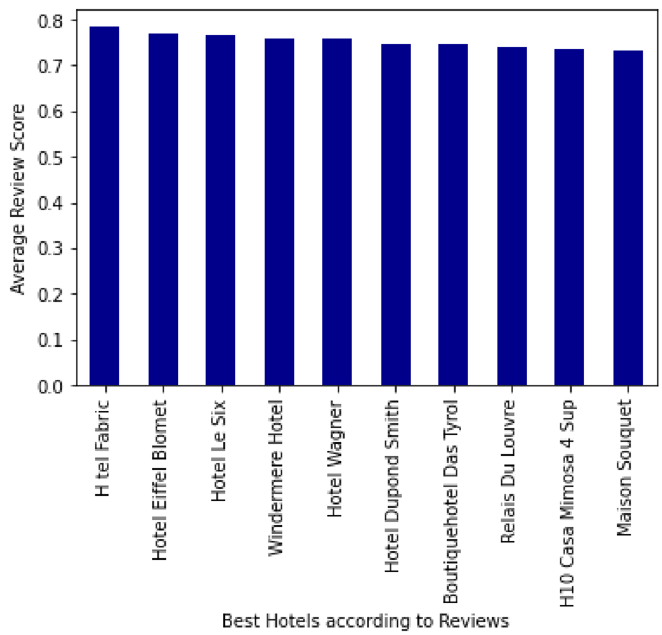

Figure 8: Bar chart to show the distribution of best hotels based on reviews

The above figure shows the best hotels based on the reviews of the customers. Hotel Fabric is the best hotel based on study.

The Worst 10 hotels according to the reviews:

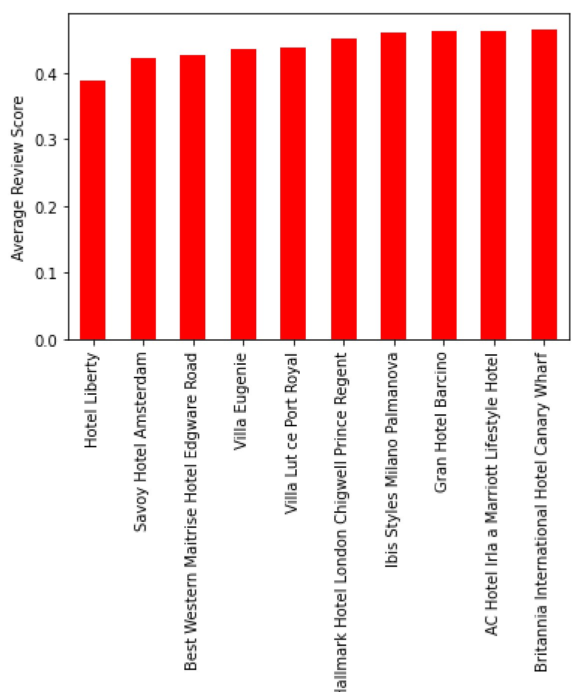

Figure 9: Bar chart to show the distribution of worst hotels based on reviews

The above figure shows the distribution of worst hotels based on the guest reviews. From the chart it can be seen that Hotel Liberty is the worst hotel and Hotel Fabric is the best hotel among 1493 hotels based on the guest reviews.

5. DISCUSSION

In this section, the findings of the study are discussed in relation to the research question. The results are discussed in detail in this part. There are a number of findings in this study which are discussed sequentially in this section. The main objective of the study is to classify the reviews into further sentiment classes so that the actual scenario of the reviews can be understood using machine learning and deep learning methods. In order to proceed with the study, extreme positive is rated as 1 and extreme negative is rated as 0. So, if a review is extremely negative, the score will be near 0 and if a review is very positive, the score will be close to 1. After cleaning the text and making it suitable for splitting the data into training and testing sets. The multilayer perceptron is used to score each review so that the intensity of the review can be understood.

5.1. Topic Modeling

Based on topic modeling, the topic which the reviewers mentioned most is "room", through this it can be evaluated that guests vent out most of their negative emotions when they are not satisfied or happy with the room they stayed. This is very clear that the rooms will be the most important factor to consider for a customer as the room is what they are mainly paying for the stay. As a result, above all hotels should invest more and pay the highest attention to the rooms that they are offering to the guests. The next interesting topic identified is "breakfast", this shows that people prefer to have a good breakfast in the hotel they spend the night and they are not happy with the breakfast menu and food or the arrangement for which it is one of the prominent topics they use when sharing negative emotions. This is something, accommodations need to be very much aware of because even after a good night's stay, a bad breakfast can make the guest's mood bad, and eventually they will search for alternatives with a good breakfast menu. Resulting in the hotel having a bad brand perception and eventually losing customers. The next topic that can be observed is "staff", efficient and friendly staff with good behavior can make a lot of difference. People love to be in those places where they get valued more and want to be with people with whom they are comfortable more. Hence, a good team of staff members can create a good environment where guests would love to interact without any hesitation and can have the best experience. The accommodations should train the staff appropriately so that they understand how to handle different customers from different cultures and backgrounds. Also, they should provide prompt and top-notch room service. Another topic that the hotels should be careful about is “bathroom”, according to the analysis, people are complaining a lot about the bathrooms they use. A bathroom is a very essential element that a guest looks for in a hotel, it should be clean with proper cleaning materials available. A bad bathroom experience can cause a customer to churn forever even after giving the best hotel room. This is highly important for any hotel to take care of and ensure hygiene and cleanliness.

5.2. Confusion Matrix

Different models are trained using the dataset in order to observe the accuracy and recall values of the models to see which model is appropriate for the dataset in predicting accurately. The properties of a classification report are important to understand which classifier is the best fit for the particular data. Precision is the ratio of true positives to the sum of true and false positives, recall is the ratio of true positives to the sum of true and false negatives, whereas F1 score is the weighted harmonic mean of precision and recall, the closer the value of F1 to 1, the better the performance of the model (Jayaswal, 2020). When the SVM classifier model is trained, the accuracy of using this classifier is 55% which means that the model is unable to predict the data accurately, only 50% of the dataset is predicted correctly. The next classifier used for training the dataset is Random Forest, Random Forest is used because it is one of the most popular machine learning classification algorithms which can deal with high dimensional noisy data. This is used here because the dataset consists of a number of features for which this can predict well with this data. Moreover, Naïve Bayes is also used to train the data because of its property of calculating probabilities which is a text classification-friendly algorithm.

Finally, the Convolutional Neural network (CNN) model is used to see the accuracy of the dataset and the accuracy is found to be 90% which is the highest among the other models used. CNN is a neural network and operates over a volume of inputs, while each layer of CNN operates to identify a pattern or useful information of the data. Convolution creates a third relationship by using a mathematical combination of two relationships (choubey, 2020). This is selected as the ideal model because even though there is a possibility of overfitting because of such high accuracy. Since this is a text classification problem, the impact of overfitting is not that high. However, the model can rightly predict True Negative (TN) and True Positive (TP) for which this model is appropriate for the dataset. From the confusion matrix it can be understood that a small amount of the data is misclassified when CNN is used for modeling.

When we do neural network, the prediction achieved is very good for which we compromise the overfitting impact because neural network can predict better than other machine learning models and also the other features in the dataset are not considered because the focus of this study was to have an in-depth accurate sentiment analysis of the reviews for which the other variables in the data are not considered. we end up having a really good prediction, we compromise with interpret because, using neural network and traditional classifiers, predict test set, which features are driving, descriptive, pattern, location, specific places bad scores, export all bad reviews

The values of the Confusion Matrix are:

True Positive: 274

True Negative: 227

False Positive: 26

False Negative: 30

		Predicted	
		Negative (**N**) -	Positive (**P**) +
Actual	Negative -	True Negatives (T**N**)	False Positives (F**P**) **Type I error**
	Positive +	False Negatives (F**N**) **Type II error**	True Positives (T**P**)

Figure 10: Confusion Matrix

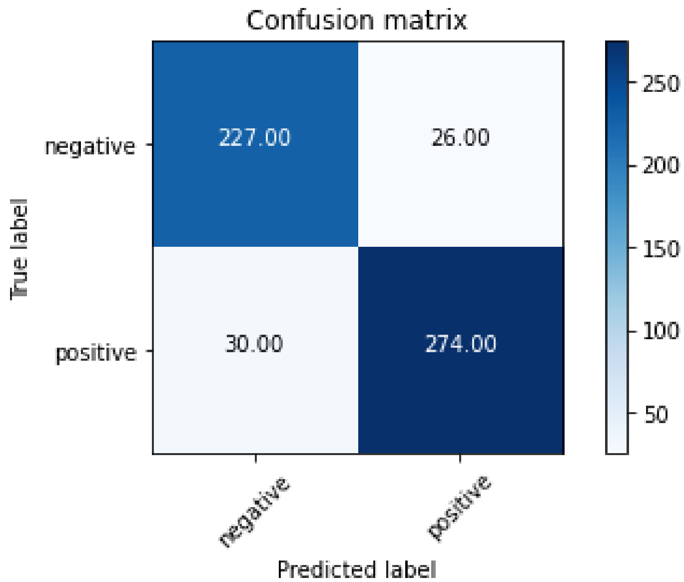

Figure 12: Confusion Matrix using CNN model

From the confusion matrix of CNN model, it can be observed that the values of true positive and true negative are high which means that the model is predicting properly mostly. Here true positive means predicted positive and it is true while true negative indicates predicted negative and it is also true. This gives a good score in recall value and precision value. These high scores show that the model is predicting well and it is suitable for the dataset. The value of True Negative (TN) and True Positive (TP) is highest for CNN compared to the other models which means that it can detect the negative and positive reviews better than any other models for which this model is suitable after doing model evaluation.

5.3. The different layers of CNN

The different layers of CNN used in this study and their purpose is described below:

Dense Layer: Dense layer is a layer which is also called as a fully connected layer which is used in the final stages of a neural network. This layer aids in reducing the dimensionality of the output from the layer before, allowing the model to more readily establish the relationship between the values of the data it is working with. It is called a connected layer because the neurons of this layer are fully connected with the neurons of the previous layer (Heidenreich, 2019). Dense layer is very common layer which is used artificial neural networks. The dense layer performs matrix vector multiplication after receiving the output from the preceding layer. An activation function is used which is responsible for the transformation of the input values of neurons. Sigmoid activation function is used for the dense layer in this study (VERMA, 2021).

Flatten Layer: This layer is a simple layer which is used to vectorize the data which is transforming the vector columns into one vector column. Here in this dataset, there are a number of columns of data, using flatten layer, all the columns are vectorized to one single column to prepare it for the neural network (ML, n.d.). The 2D output from CNN model is flattened to one long 2D vector to represent the features.

Embedding layer: An input is transformed from a sparse representation into a distributed or dense representation by the embedding layer of a neural network. It is used in this study because it is a common use in Natural Language Processing to take a sparse word and then make it dense. The embedding layers are trained a number of times using Word2Vec (Saxena, 2020).

Conv-1D layer: This is the layer that is used for creating the convolution kernel, this layer is coiled with the input layer and produces tensor of outputs (Sharma, 2020). In Conv-1D only one dimension is used for which the convolution operates only in the first axis.

GlobalMaxPooling1D layer: It is a layer that calculates the maximum value for patches of a feature map, and uses it to create a downsampled (pooled) feature map (satyam00so, 2022). This layer is used after a convolution layer.

5.4. Sentiment scoring

Using the Multilayer perceptron model, all the text reviews are scored which are used as features in the analysis. Based on the score the emotions reflected in the text is measured. This will help to understand the customer's feelings towards the brand. This can directly help the company to understand the brand perception of the consumers. The study involves customer sentiment analysis which is the process of emotion mining to know exactly what the consumers have in mind about the brand. In this study, the reviews are scored between 1 and 0, and then it is observed that most of the reviewers added both positive and negative reviews for which the scores are difficult to identify.

Hence, to solve this issue, a final average sentiment score is calculated which is the average of the scores of positive and negative predictions. This score is mentioned as a Sentiment score in the code. Based on the sentiment score the reviews are divided into 4 classes. The classes are produced in this way, reviews scoring higher than 0.7 are considered as "best", reviews scored in between 0.7 and 0.5 are labelled as “good”. “Bad” reviews are those with scores in between 0.5 and 0.3 and the rest below 0.3 are labelled as “worst”. By evaluating the emotions from reviews, it can be observed that the majority of the listed hotels on the Booking.com website is considered good from the consumer’s perspective. This indicates that the consumers of Booking.com have a positive brand perception as most of them are expressing positive emotions in reviews. From the final visualization, this can be stated that the people are supporting the hotels by adding suggestive and advising comments in the reviews and the number of reviewers leaving a positive review only is higher than those leaving a negative review only.

In the final visualization of the scores, it can be stated that a number of negative reviews are not totally negative, it can be assumed that reviewers might have added some advices along with a positive review which appeared as a score for negative review. In addition, it can be observed that the number of guests leaving a positive review alone is higher than the guests leaving just a negative review. Since, most of the reviewers also stated a negative review with positive one, an average sentiment score is calculated which is the average of positive and negative prediction. This is the score based on which the reviews are divided into 4 classes from Best to Worst.

Finally, the hotels with the highest average review score and the lowest average review score are evaluated and visualized. This will help the company to identify the top performing hotels and below average performing hotels based on which marketing plans can be developed to leverage the low performers. Among the low performers, the remarkable ones are Hotel Liberty, Savoy Hotel Amsterdam, etc. Also, the top performers include Hotel Fabric, Hotel Eiffel Blomet, etc. Further studies can be performed on the best and worst ones and then compared to understand the impactful factors.

The objective of the research is to perform a sentiment analysis on the hotel reviews extracted from Booking.com website in order to classify the textual reviews into sentiment classes and to identify the main reasons behind the negative reviews. From the analysis, the areas of focus for the hotels has been discovered and also the reviews are further classified into classes to understand the emotions of the customers and the best and worst hotels in Europe is distinguished.

6. CONCLUSION & RECOMMENDATION

In this study, a neural network model called CNN is applied to perform the sentiment analysis of hotel reviews extracted from the Booking.com website. From the outcome of the analysis it can be claimed that well-trained CNN model can perform better than traditional machine learning algorithms for sentiment classification. In this project, the reviews are labeled into 4 different classes such as Best, Good, Bad and Worst. Since, there was a problem that many reviewers have given both positive and negative reviews, an average sentiment score is calculated and then based on that score the classes are divided. The accuracy of the CNN model is found to be 90% which is very high compared to the machine learning models such as SVM, Random Forest, and Naïve Bayes classification models. Topic modeling is also applied to the negative reviews of the dataset and the important topics identified are "room", "bathroom", "breakfast", "bed", etc. Finally, the best and worst hotels are also listed according to the outcome of the scores of the reviews. Initially, the dataset is preprocessed by using tokenization and other cleaning methods like lowering the case of the words and removing punctuations and unwanted signs and symbols. This is done so that the accuracy can be performed accurately. A vocabulary of the words is created because an embedding model is used. Embedding is executed in this study using the Word2Vec model and the train and test size is split into 67% training size and 33% test size for getting higher accuracy of the model. CNN is used in this analysis because CNN can better predict text classification problems.

One recommendation for the company would be to work on these best and worst hotels. Booking.com can give suggestions and advices to the worst hotels so that they can be improved and some promotional plans can be executed to prevent them from falling behind. On the other hand, the best hotels should be awarded and some special offers and announcements should be made on the website and application of Booking.com so that the brand perception of the hotels can be improved. This study is performed with the objective to make an in-depth sentiment analysis of the reviews of Booking.com so that both the consumers and the hotels can have a strong relationship by enhancing the brand perception. To reduce the gap between the expectation of guests and the amenities offered by a hotel. Another recommendation is that Booking.com should create a team to investigate the topics associated with the negative reviews and communicate with the clients so that the hotels can improve the areas guests are more concerned about. These hotels are in high demand as Europe is one of the best and most beautiful tourist destinations for people all over the world (Jess, 2015). Hence, it is very much important to give a wonderful stay experience to the visitors so that more people come to visit and enjoy the beauty of European countries.

One limitation of the research can be the use of CNN model with only one parameter setting. Also, the dataset considered in this research only covers hotels in European countries, the result or outcome cannot be considered to assume the scenario of other parts of the country. Future work includes experimentation with other deep learning models and by changing the parameter settings of the CNN model and comparing the results. In addition, different datasets can be used for other industries like fast-moving consumer goods industry product reviews, restaurant reviews from other review sites, movie reviews, and online e-commerce site reviews. The reviews can be performed using this technique to get a much better and clear image of the brand perception of a company or organization.

7. REFERENCES

Abirami, S. & Chitra, P., 2020. The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases. *ScienceDirect.*

Asghar, N., 2016. Yelp Dataset Challenge: Review Rating Prediction. *University of Waterloo.*

Bento, C., 2021. *Towards Data Science.* [Online]
Available at: https://towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141
[Accessed 02 09 2022].

Bernardi, L., Mavridis, T. & Estevez, P., 2019. 150 Successful Machine Learning Models: 6 Lessons Learned at Booking.com. *Applied data Scinece track paper.*

Borges-Tiago, M. T., Arruda, C., Tiago, F. & Rita, P., 2021. Differences between TripAdvisor and Booking.com in branding co-creation. *Journal of Business Research,* pp. 380-388.

choubey, v., 2020. *Medium.* [Online]
Available at: https://medium.com/voice-tech-podcast/text-classification-using-cnn-9ade8155dfb9
[Accessed 23 August 2022].

Datafloq, 2014. *SmartDataCollective.* [Online]
Available at: https://www.smartdatacollective.com/why-hotels-should-apply-big-data-analytics-provide-unique-guest-experience/
[Accessed 20 July 2022].

Euromonitor International, n.d. [Online]
Available at: http://go.euromonitor.com/rs/805-KOK-719/images/2017%20Top%20100%20Cities%20Destinations%20Final%20Report.pdf (2017)

Fuentes, E. M., 2016. Are guests of the same opinion as the hotel star-rate classification system?. *Journal of Hospitality and Tourism Management,* pp. 126-134.

Gandhi, R., 2018. *Towards Data Science.* [Online]
Available at: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c
[Accessed 25 08 2022].

Habimana, O. et al., 2020. Sentiment analysis using deep learning approaches: an overview. *Science China Information Sciences,* Issue 2019, p. 63.

Harrison-Walker, L. J., 2001. "E-complaining: a content analysis of an Internet complaint forum. *Journal of Services Marketing,* Volume 15.

Heidenreich, H., 2019. *Medium.* [Online]
Available at: https://medium.com/p/2abadff9b990
[Accessed 26 08 2022].

Jayaswal, V., 2020. *Towards Data Science.* [Online]
Available at: https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262#:~:text=Confusion%20matrix%2C%20precision%2C%20recall%2C%20and%20F1%20score%20provides,segmentation%2C%20named%20entity%20recognition%2C%20and%2
[Accessed 31 08 2022].

Jess, 2015. *The Travelista.* [Online]
Available at: https://thetravelista.net/2015/03/09/9-reasons-why-europe-is-the-best-continent-in-the-world-and-why-you-should-feel-lucky-to-live-here/
[Accessed 02 09 2022].

Khotimah , D. A. K. & Riyanarto , S., 2018. Sentiment Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis. *2018 6th International Conference on Information and Communication Technology (ICoICT).*

Kulshrestha, R., 2019. *Towards Data Science.* [Online]
Available at: https://towardsdatascience.com/nlp-101-word2vec-skip-gram-and-cbow-93512ee24314
[Accessed 26 08 2022].

Lee, H. & Blum, S. C., 2015. How hotel responses to online reviews differ by hotel rating: an exploratory study. *Worldwide Hospitality and Tourism Themes.*

Li, S., 2018. *Towards Data Science.* [Online]
Available at: https://medium.com/p/9bf156893c24
[Accessed 01 09 2022].

M. Ady, &. D. Q.-F., 2015. Consumer research identifies how to present travel review content for more bookings. *IEEE,* p. 1.

M. Mariani, M. & Borghi, M., 2017. Effects of the Booking.com rating system: Bringing hotel class into the picture. *ELSEVIER,* pp. 48-50.

Manthiramoorthi, M., n.d. *OpenGenus.* [Online]
Available at: https://iq.opengenus.org/topic-modeling-nmf/
[Accessed 30 08 2022].

ML, V., 2017. *Value ML.* [Online]
Available at: https://valueml.com/keras-flatten-operation-in-cnn-models-in-machine-learning/
[Accessed 25 08 2022].

ML, V., n.d. *Value ML.* [Online]
Available at: https://valueml.com/keras-flatten-operation-in-cnn-models-in-machine-learning/
[Accessed 23 8 2022].

Mumbi, W., 2021. *Section.* [Online]
Available at: https://www.section.io/engineering-education/what-is-word2vec/
[Accessed 24 08 2022].

Patel, H., 2018. *OpenGenus IQ.* [Online]
Available at: https://iq.opengenus.org/author/harshiv/
[Accessed 01 09 2022].

Perikos, I. et al., 2018. Opinion Mining and Visualization of Online Users Reviews : A Case Study in Booking.com. *IEEE.*

Rani, S. & Kumar, P., 2019. Deep Learning Based Sentiment Analysis Using Convolution Neural. *Arabian Journal for Science and Engineering,* Issue 2018, pp. 3305-3314.

Rodrı´guez Dı´az, M. & F Espino Rodrı´guez, T., 2018. Determining the reliability and validity of online reputation databases for lodging: Booking.com, TripAdvisor, and HolidayCheck. *Journal of Vacation Marketing,* Volume 24(3), pp. 261-274.

satyam00so, 2022. *Geeks for Geeks.* [Online]
Available at: https://www.geeksforgeeks.org/tensorflow-js-tf-layers-globalmaxpooling1d-function/
[Accessed 01 09 2022].

Saxena, S., 2020. *Medium.* [Online]
Available at: https://medium.com/analytics-vidhya/understanding-embedding-layer-in-keras-bbe3ff1327ce
[Accessed 23 8 2022].

Sharma, P., 2020. *Machine Learning Knowledge.* [Online]
Available at: https://machinelearningknowledge.ai/keras-convolution-layer-a-beginners-guide/#:~:text=The%20Conv-1D%20Layer%20of%20Keras%20is%20used%20for,added%20to%20outputs%20if%20it%E2%80%99s%20passed%20as%20true.
[Accessed 24 8 2022].

SHI, H.-X. & LI, X.-J., 2011. A SENTIMENT ANALYSIS MODEL FOR HOTEL REVIEWS BASED ON SUPERVISED LEARNING. *International Conference on Machine Learning and Cybernetics.*

Singh, V. et al., 2013. Sentiment Analysis of Textual Reviews. *5th International Conference on Knowledge and Smart Technology (KST).*

Solanki, S., 2022. *CoderzColumn.* [Online]
Available at: https://coderzcolumn.com/tutorials/artificial-intelligence/keras-cnn-with-conv1d-for-text-

classification
[Accessed 01 09 2022].

Stecanella, B., 2017. *Monkey Learn.* [Online]
Available at: https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/#:~:text=A%20support%20vector%20machine%20(SVM,able%20to%20categorize%20new%20text.
[Accessed 26 08 2022].

VERMA, Y., 2021. *Analytics India Mag.* [Online]
Available at: https://analyticsindiamag.com/a-complete-understanding-of-dense-layers-in-neural-networks/#:~:text=A%20dense%20layer%20also%20referred%20to%20as%20a,the%20data%20in%20which%20the%20model%20is%20working.
[Accessed 23 08 2022].

Wang, M., Zhao, Y. & Xu, X., 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management,* Volume 76, pp. 111-121.

Xu, X., 2018. Examining The Relevance Of Online Customer Textual Reviews On Hotels' Product And Service Attributes. *Journal of Hospitality and Tourism Research,* Volume 43, pp. 141-163.

Xu, X., 2020. Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants. *ELSEVIER,* Volume 106, pp. 196-210.

Xu, X., Wang, X., Li, Y. & Haghighi, M., 2017. Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *ELSEVIER,* Volume 37, pp. 673-683.

Yiu, T., 2019. *Towards Data Science.* [Online]
Available at: https://towardsdatascience.com/understanding-random-forest-58381e0602d2
[Accessed 26 08 2022].

Zhang, X. & Kim, H.-S., 2021. Customer Experience and Satisfaction of Disneyland Hotel through Big Data Analysis of Online Customer Reviews. *Sustainability,* Volume 13.